

The Data Scientist's Toolbox

(online short course - 10 hrs)

Lecturer: Balázs Laki

E-mail: laki.balazs@edutus.hu

Topics:

- *Data Science Fundamentals*
 - What is Data Science?
 - What is Data?
 - Getting Help
 - The Data Science Process
- *R and RStudio*
 - Installing R
 - Installing R Studio
 - RStudio Tour
 - R Packages
 - Projects in R
- *Version Control and GitHub*
 - Version Control
 - Github and Git
 - Linking Github and R Studio
 - Projects under Version Control
- *R Markdown, Scientific Thinking, and Big Data*
 - R Markdown
 - Types of Data Science Questions
 - Experimental Design
 - Big Data
- *Exploratory Data Analysis*
 - Principles of Analytic Graphics
 - Exploratory Graphs
 - Plotting Systems in R
 - Base Plotting System (part 1)
 - Base Plotting Demonstration
 - Graphics Devices in R
 - Lattice Plotting System
 - ggplot2
 - Hierarchical Clustering
 - K-Means Clustering
 - Dimension Reduction
 - Working with Color in R Plots
 - Clustering Case Study
 - Air Pollution Case Study